

BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications

Patricia L. Whetzel*, Natalya F. Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache and Mark A. Musen

Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA 94305, USA

Received March 18, 2011; Revised May 13, 2011; Accepted May 21, 2011

ABSTRACT

The National Center for Biomedical Ontology (NCBO) is one of the National Centers for Biomedical Computing funded under the NIH Roadmap Initiative. Contributing to the national computing infrastructure, NCBO has developed BioPortal, a web portal that provides access to a library of biomedical ontologies and terminologies (<http://bioportal.bioontology.org>) via the NCBO Web services. BioPortal enables community participation in the evaluation and evolution of ontology content by providing features to add mappings between terms, to add comments linked to specific ontology terms and to provide ontology reviews. The NCBO Web services (http://www.bioontology.org/wiki/index.php/NCBO_REST_services) enable this functionality and provide a uniform mechanism to access ontologies from a variety of knowledge representation formats, such as Web Ontology Language (OWL) and Open Biological and Biomedical Ontologies (OBO) format. The Web services provide multi-layered access to the ontology content, from getting all terms in an ontology to retrieving metadata about a term. Users can easily incorporate the NCBO Web services into software applications to generate semantically aware applications and to facilitate structured data collection.

INTRODUCTION

Ontologies provide domain knowledge to drive data annotation, data integration, information retrieval, natural language processing and decision support. As the number of large data sets are growing, providing a framework for data analysis and data integration using ontologies

continues to be of critical importance (1). However, until recently, there has been a lack of common services for accessing this rich content from software applications. There has also been a lack of services to facilitate ontology development by reusing existing ontology content. BioPortal fills these gaps (2). BioPortal is a Web portal that provides access to a library of biomedical ontologies and terminologies developed in Web Ontology Language (OWL), Resource Description Framework (RDF)(S), Open Biological and Biomedical Ontologies (OBO) format, Protégé frames and Rich Release Format (<http://bioportal.bioontology.org>). BioPortal has a service-oriented architecture; the NCBO Web services provide the functionality found in BioPortal and these Web services can be incorporated into other software applications to access and use ontology content. BioPortal groups ontologies by domain to ease finding relevant ontologies and allows users to browse, search and visualize the content of ontologies. Registered users are able to add mappings between terms, to add comments on individual terms within the ontology and to provide reviews of ontologies. This user-generated content provides critical evaluation and feedback mechanism for ontology developers. The specific focus on enabling community feedback to BioPortal content is a distinguishing characteristic of the system.

ONTOLOGY DATA

In 2008, BioPortal contained 72 ontologies (300 000 total classes) and has grown significantly over the last 3 years to contain 260 ontologies (4.8 million total classes). Ontologies from a number of different groups are published in BioPortal, including caBIG (<https://cabig.nci.nih.gov/>), the recipients of Clinical and Translational Science Awards (<http://www.ctsaweb.org/>), the Consultative Group on International Agricultural Research (<http://www>

*To whom correspondence should be addressed. Tel: +650 721 2378; Fax: +650 725 7944; Email: whetzel@stanford.edu

.cgiar.org/), the OBO library (<http://obofoundry.org/>), the Proteomics Standards Initiative (<http://www.psidev.info/>), the Unified Medical Language System (<http://www.nlm.nih.gov/research/umls/>) and the World Health Organization (WHO) Family of International Classifications (<http://www.who.int/classifications/en/>). In addition to the increase in ontology content within BioPortal, non-biomedical organizations have also installed their own instances of BioPortal software. These organizations include DataONE (<http://www.dataone.org/>), the Marine Metadata Interoperability Project (<http://mmisw.org/orr/>) and other groups that require Official Use Only levels of privacy for their ontology content and access to the Web services (e.g. annotating HIPPA regulated data) or need a repository for ontologies that cover domains not relevant to biomedicine.

WEB SERVICES

When we initially released BioPortal, the system included RESTful Web services to get ontology metadata, to get individual ontology terms, to download ontologies and to search within ontologies. Since then, we have increased the number of Web services to provide expanded functionality and to include Web services to create and get ontology views, to get all terms from an ontology, to get instances, to post and get ontology mappings, to post and get comments and to get ontologies and individual ontology terms in RDF. BioPortal is designed to store multiple versions of the same ontology, which enables a historical overview of the ontology as it evolves. Each ontology has a global

(virtual) ontology identifier and each new version of the ontology has an ontology version identifier. Many of the Web services can be called with either the virtual ontology identifier or the ontology version identifier.

Ontology views are subsets of one or more ontologies. Ontology subsets are also referred to as slims by the GO Consortium and value sets when used for structured data entry. These subsets are a useful mechanism to work with smaller amounts of ontology content. For example, views can serve as value sets to populate a Web form select menu or as portions of ontologies to re-use in developing a new ontology. The 'View' Web services include functionality to get a list of all ontologies that have views and to create a view using the 'View extraction' Web service. The View extraction Web service is designed to extract branches of ontologies given a term to serve as the root node in the ontology view. This Web service is very popular for generating views of content specific portions of large ontologies such as the NCBI Taxonomy, International Classification of External Causes of Injuries (<http://www.who.int/classifications/icd/adaptations/iceci/en/index.html>) and the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT, <http://www.ihstso.org/snomed-ct>).

The 'Get All Terms' Web service returns the term details for all terms within an ontology. This Web service can be called using a specific ontology version identifier or with the virtual ontology identifier, therefore providing a common Web service signature that always returns data from the latest version of the ontology. Due to the large size of some ontologies, this Web service returns 'pages' of

The screenshot shows the 'Biomedical Resource Ontology Version 3.1' interface. The 'Physiology Facility' is selected, with a 'Link Here' and 'Subscribe' button. The 'View Ontology Summary' tab is active, showing a tree view of the ontology structure. The 'Details' tab is also visible. A table of 'Notes (3)' is displayed, with columns for 'SUBJECT', 'AUTHOR', 'TYPE', and 'CREATED'. The table contains three entries related to 'New Term Proposal: Sleep Study Facility' and 'New Term Proposal: Exercise Study Facility'. Below the table, there is a 'Hide Add Note' button and a 'New Term' form. The form includes fields for 'Preferred name', 'Synonyms (comma-separated list)', 'Definition', 'Proposed id', 'Parent (start typing to autocomplete)', and 'Contact information (publicly visible)'. A 'submit' button is at the bottom.

SUBJECT	AUTHOR	TYPE	CREATED
Review definition	whetzel	Comment	03/14/2010
New Term Proposal: Sleep Study Facility	Mette	New Term Proposal	06/15/2010
New Term Proposal: Exercise Study Facility	Mette	New Term Proposal	06/15/2010

Showing 1 to 3 of 3 entries
[First](#) [Previous](#) [1](#) [Next](#) [Last](#)

New Term

Reason for change *

Preferred name *

Synonyms (comma-separated list)

Definition *

Proposed id

Parent (start typing to autocomplete) *

<http://bioontology.org/ontologies/BiomedicalResourceOnto>

Contact information (publicly visible)

submit

* required

Figure 1. Screenshot of New Term proposal BioPortal Note Type.

Figure 2. BioPortal Widgets. Widgets are available for all ontologies in BioPortal. To view a demo of the Widget and get the code to use in your web site, click on the 'Ontology Widgets' tab.

data to minimize the load on both client and server in dealing with extremely large XML files. The Web service is particularly valuable for use of ontology data in other knowledge base systems that require a custom ontology format, such as the Ontology Management Cell from the i2b2 clinical research data management Hive (<https://www.i2b2.org/software/index.html>).

The 'Get Term' Web service is now expanded to access instances from OWL ontologies. The Web service returns 'pages' of results containing all instances for a class. Based on the design of the ontology, it is useful to access the instances since these may be the terms to use for data annotation. For example, in the MGED Ontology the values to specify a requested Minimum Information About a Microarray Experiment (MIAME) checklist item are located as instances in the ontology.

The suite of 'Mapping' Web services provides access to the millions of ontology mappings published in BioPortal. The mapping data includes mappings provided by the ontology content providers, for example, mappings based on common Concept Unique Identifiers (CUI) in UMLS, mappings specified in OBO ontologies through the OBO xref property, mappings submitted directly to BioPortal by users and mappings generated automatically by algorithms such as LOOM (3). The Mapping Web services are parameterized to allow a high degree of flexibility to access the data. For example, the service can return mappings between individual terms, all mappings for a given term, or all mappings for a given ontology. Registered users can

also submit mappings directly to BioPortal by using the 'create new mapping' Web service. This service allows automatic publishing of mapping content generated by ontology alignment software.

The 'Notes' Web services provide the ability for registered users to add comments directly to an ontology term or to the ontology. There are different types of notes that provide varying levels of structure for the note. For example, the note type 'comment' simply provides a text box to add unstructured comments. However, the 'new term proposal' note type provides fields to include the preferred name, synonyms, definition and reason for proposing the new term (Figure 1). Each note type has a specific XML response from the Notes Web service, easily allowing software applications to post or consume the content of the Notes Web service. For example, via the service-oriented architecture of BioPortal, the notes functionality is being incorporated into the ontology editing software WebProtégé (<http://protegewiki.stanford.edu/wiki/WebProtege>) and could also be incorporated into OBO-Edit (<http://oboedit.org/>). An email alert system is incorporated into the Notes feature, therefore the ontology author and other users interested in comments for a given ontology can be notified via email of new comments. The ontology author can then engage with the user to address the concern using the Notes feature and archive the Note once resolved.

The 'RDF' Web services are designed to return RDF snippets for individual terms and the content of an entire

Table 1. Software applications semantically enabled by NCBO Web services and Widgets

Software	NCBO Technology	URL
ALEX	W	https://alex.med.nyu.edu/portal/
aTag Generator	W	http://hcls.deri.org/ataq/generator/
BioLit	O	http://biolit.ucsd.edu/doc/
BioPortal Reference Plugin	O	http://protege.wiki.stanford.edu/wiki/BioPortal_Reference_Plugin
BioPortal Import Plugin	O	http://protege.wiki.stanford.edu/wiki/BioPortal_Import_Plugin
BioScholar	O	https://wiki.birncommunity.org/display/NEWBIRNCC/BioScholar
cgMDR	O	http://goo.gl/dvsgM
CISBIC Data Management	O	http://www3.imperial.ac.uk/cisbic/corefacilities/datamanagement
eleMAP	O	https://vict.vanderbilt.edu/eleMAP/
GeneWiki	A	http://en.wikipedia.org/wiki/Portal:Gene_Wiki
iCAT	O	http://sites.google.com/site/icd11revision/home/icat
ISAcreeator	O, A	http://isatab.sourceforge.net/isacreeator.html
Jinx	O	http://ncmir.ucsd.edu/downloads/jinx.shtm
Knowledge Egg	W	http://www.kunnskapsegget.no/
Microsoft Word Addin for Ontology Recognition	O	http://ucsdbiolit.codeplex.com/
NEMO Toolkit	O	http://nemo.nic.uoregon.edu/wiki/NEMO_ERP_Analysis_Toolkit
Nutritional Phenotype database	W	http://www.dbnp.org/
ODIE	A	http://www.bioontology.org/ODIE-project
ODiSSea	O, A, RI	http://www.hub.sciverse.com/
OntoCat	O	http://www.ontocat.org/
Ontological Discovery Environment	A	http://ontologicaldiscovery.org/
openMDR	O	http://citih.osumc.edu/projects/project&r=1032
Oryzabase	W	http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp
RadLex Tree Browser	O, W	http://www.radlex.org/
RadSpeech	W	http://digitaleveredelung.dfki.de/MEDICO-Playground/term2.html
REDfly	W	http://redfly.ccr.buffalo.edu/
RightField	O	http://www.sysmo-db.org/rightfield
S3DB	O	https://sites.google.com/a/s3db.org/s3db/
SEEK	O	http://www.sysmo-db.org/seek
Semantic Medical Image Annotation	O	http://kithira.biosim.ntua.gr/semia
Sentient Knowledge Explorer	O	http://www.io-informatics.com/
SimTK	W	https://simtk.org/home/simtk
STRIDE	A	https://clinicalinformatics.stanford.edu/research/stride.html
TRIAD	O	http://triadcommunity.org/
Tripod	A	http://tripod.nih.gov/
Zooma	O	http://zooma.sourceforge.net/

A: Annotator Web service, O: Ontology Web services, RI: Resource Index Web service, W: Widgets.

ontology in RDF. The goal of the RDF Web services is to provide the essential information about a term in RDF. The RDF Term Web service returns the term id, preferred name, synonyms, definitions and super-classes together with selected locally defined annotation properties for the ontology. The URI for each term is either the original URI present in the ontology file or the URI specified by the respective authority. If the ontology developers do not provide URIs for the terms, BioPortal generates these using the purl.bioontology.org server. By generating RDF for the ontologies, the entire content of BioPortal ontologies can be exposed as Linked Open Data. In addition to the RDF Web services, the BioPortal prototype SPARQL endpoint provides access to BioPortal ontology content in RDF and is available at: <http://sparql.bioontology.org>.

We have also developed other Web services that use BioPortal ontology content. These services include the NCBO Annotator Web service, which ‘tags’ textual data with ontology terms from BioPortal (4); the NCBO Resource Index Web services, which provide access to an ontology-based index of publically available

biomedical data (5) and the NCBO Ontology Recommender Web service (6), which given a set of keywords or textual data as input generates a ranked score of which ontology best ‘covers’ the data. More details on all NCBO Web services can be found at: http://www.bioontology.org/wiki/index.php/NCBO_REST_services.

WIDGETS

We have also wrapped the NCBO Web services as widgets for easy embedding of this functionality into Web sites. The ‘Form Autocomplete’ widget provides variations to fill-in the form menu with the full term URI, term identifier, or the term name (Figure 2). The ‘Jump To’ widget also provides the term autocomplete function, but allows users to ‘jump’ directly into BioPortal to learn more about the term (e.g. term details, position in hierarchy, database records annotated with the term). The ‘RSS Feed’ provides updates on new versions and comments posted to the ontology and can be added to Web sites to notify community members of these updates. The ‘Graph’ and

'Tree' widgets provide a way to display an ontology structure on your own web site. The Widgets are available for each ontology in BioPortal and the widget code is accessible from the Widgets tab.

DISCUSSION

BioPortal is a Web portal that provides access to a library of biomedical ontologies and terminologies via the NCBO Web services. While the overall architecture is domain independent, the NCBO instance is focused on publishing ontology content relevant for biomedicine. The ontology content in BioPortal continues to grow and the usage of the site continues to increase. Over the last 2 years, the number of ontology terms in BioPortal has grown by more than an order of magnitude and the number of ontologies increased 4-fold. The usage data for 2010 shows that almost 36 000 unique visitors accessed 35 million pages via the BioPortal Web services, processing ~100 GB of data.

The ease of use of the NCBO Web services and Widgets provide a convenient mechanism to incorporate ontology content into software applications. A number of data-annotation applications use NCBO Web services, including ISAcreeator (7) for annotation of high-throughput experimental data, openMDR (8) a caBIG tool for annotation of data elements relevant to cancer related studies, the ECG Gadget (<http://wiki.cvrgrid.org/index.php/ECGGadget>) for annotation of electrocardiographs and BioScholar (<https://wiki.birncommunity.org/display/NEWBIRNCC/BioScholar>) a tool from BIRN to construct a knowledge repository derived from PDF files. The Widgets have also been added to a number of Web sites to enhance data annotation. These Web sites include REDfly, a database of *Drosophila* transcriptional *cis*-regulatory elements (9), the RadLex Tree Browser that provides a radiologist-friendly view of the RadLex terminology (10,11) and the model organism databases such as Oryzabase (12). The NCBO Web services are facilitating ontology-based search of content in caBIG resources and ODisSea from Elsevier that integrates journal publications in SciVerse Hub with publically available databases via an ontology-based index. Table 1 lists additional software applications using NCBO technology. Additional use cases and sample code can be accessed from: http://www.bioontology.org/wiki/index.php/Sample_Code_Cookbook.

NCBO is continuing to develop new ontology-based tools. The User Group (<https://mailman.stanford.edu/mailman/listinfo/bioportal-user-group>) provides a mechanism to learn more about upcoming features and provide feedback. User feedback can also be sent to our support mailing list (support@bioontology.org).

ACKNOWLEDGEMENTS

We thank Alex Skrenchuk from Stanford University for computer support.

FUNDING

National Center for Biomedical Ontology, under roadmap-initiative from the National Institutes of Health [grant U54 HG004028]. Funding for open access charge: National Institutes of Health [grant U54 HG004028].

Conflict of interest statement. None declared.

REFERENCES

- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre, S. *et al.* (2008) Big data: the future of biocuration. *Nature*, **455**, 47–50.
- Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G. *et al.* (2009) BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.*, **37**, W170–W173.
- Ghazvinian, A., Noy, N.F. and Musen, M.A. (2009) Creating mappings for ontologies in biomedicine: simple methods work. *AMIA Annu. Symp. proc.*, **2009**, 198–202.
- Jonquet, C., Shah, N.H. and Musen, M.A. (2009) The open biomedical annotator, San Francisco. *Summit Trans Bioinformatics*, **2009**, 56–60.
- Shah, N.H., Jonquet, C., Chiang, A.P., Butte, A.J., Chen, R. and Musen, M.A. (2009) Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinformatics*, **10**(Suppl 2), S1.
- Jonquet, C., Musen, M.A. and Shah, N.H. (2010) Building a biomedical ontology recommender web service. *J. Biomed. Sem.*, **1**(Suppl 1), S1.
- Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., Hofmann, O. *et al.* (2010) ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics*, **26**, 2354–2356.
- Dhaval, R., Melean, C., Ervin, D. and Payne, P.O. (2010) *Using the Open Metadata Registry (OpenMDR) to Create Data Sharing Interfaces*. CTSA IKFC Annual Meeting, Washington, D.C.
- Gallo, S.M., Gerrard, D.T., Miner, D., Simich, M., Des Soye, B., Bergman, C.M. and Halfon, M.S. (2011) REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res.*, **39**, D118–D123.
- Langlotz, C.P. (2006) RadLex: a new method for indexing online educational materials. *Radiographics: a Review Publication of the Radiological Society of North America, Inc.*, **26**, 1595–1597.
- Kundu, S., Itkin, M., Gervais, D.A., Krishnamurthy, V.N., Wallace, M.J., Cardella, J.F., Rubin, D.L. and Langlotz, C.P. (2009) The IR RadLex project: an interventional radiology lexicon—a collaborative project of the Radiological Society of North America and the Society of Interventional Radiology. *J. Vasc. Interv. Radiol.*, **20**, 433–435.
- Kurata, N. and Yamazaki, Y. (2006) Oryzabase. an integrated biological and genome information database for rice. *Plant physiol.*, **140**, 12–17.